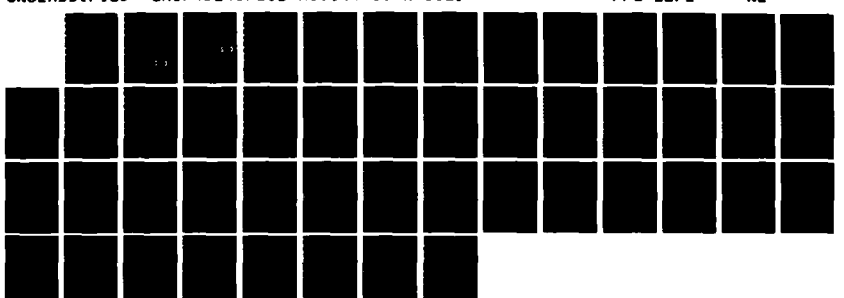
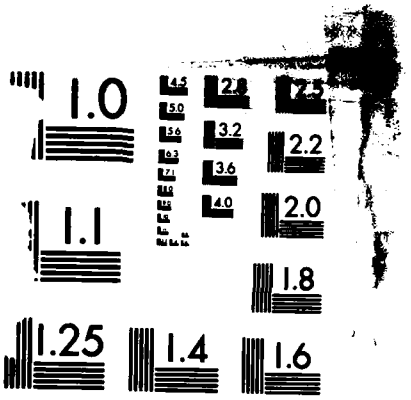
AD-A165 892 MATHEMATICAL AND STATISTICAL CHARACTERIZATIONS OF GENERALIZED HYPEREXPONE... (U) GEORGE MASON UNIV FAIRFAX VA DEPT OF SYSTEMS ENGINEERING R F BOTTA ET AL. FEB 86 1/1

UNCLASSIFIED GMU/49146/101 N00044-86-K-0029 F/G 12/1 NL





The State University
in Northern Virginia

(12)

George Mason University

MATHEMATICAL AND STATISTICAL
CHARACTERIZATIONS OF GENERALIZED HYPER-
EXPONENTIAL DISTRIBUTION FUNCTIONS

by

Robert F. Botta
Carl M. Harris
William G. Marchal

(12)

TECHNICAL REPORT

Office of Naval Research

Contract No. N00014-86-K-0029

MATHEMATICAL AND STATISTICAL CHARACTERIZATIONS OF GENERALIZED HYPER-EXPONENTIAL DISTRIBUTION FUNCTIONS

by

Robert F. Botta
Carl M. Harris
William G. Marchal

Report No. GMU/49146/101

February 1986

Department of Systems Engineering
School of Information Technology and Engineering
George Mason University
Fairfax, Virginia 22030

Copy No. 9

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. ADA 165892	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Mathematical and Statistical Characterizations of Generalized Hyperexponential Distribution Functions		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER GMU/49146/10$ /
7. AUTHOR(s) Robert F. Botta, Carl M. Harris, William G. Marchal		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0029
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Systems Engineering, George Mason University, Fairfax, VA 22030		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Task H-B, Project 4118150
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217		12. REPORT DATE February 1986
		13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE Unclassified

16. DISTRIBUTION STATEMENT (of this Report)
Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
distribution functions
exponential distributions
hyperexponential distributions
method of stages
approximations
characterizations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
This paper examines in detail the class of generalized hyperexponential (GH) probability distribution functions. The family is compared to and contrasted with similar popular classes of distributions used in stochastic modeling. Each of these families arises from a desire to preserve the computationally attractive feature of "memorylessness" possessed by the exponential probability distribution while extending the representations to a broader class in order to approximate an arbitrary probability distribution function. Thus, the simple structure and attractive properties of the GH

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

probability distribution functions are presented with a view toward facilitating the mathematical operations which frequently occur in practice.

1. Introduction

The Markovian "memoryless" property of the exponential probability distribution simplifies many conditional probability calculations. Exponential functions are also the solution to many natural elementary differential equations which model a broad range of applications. As such, the exponential distribution is a "workhorse" of applied probability.

A first attempt to generalize the exponential family was undertaken by Erlang and extended by Jensen [1954]. Erlang considered a series of independent exponential distributions all with the same parameter. Then Jensen defined the generalized Erlang (GE) family of probability distribution functions, which have the interpretation that the modeled process consists of successive independent stages each having an exponential distribution. The resultant mathematical operation is seen to be a convolution since the duration is distributed as the sum of several, not necessarily identical, exponential random variables.

The GE family increased modeling flexibility, but has the restrictive property that all GEs have a coefficient of variation less than one. To create probability distribution functions whose coefficient of variation exceeds unity, one considers mixtures in contrast to convolutions of exponential distributions. This leads to the hyperexponential family of probability distribution functions.

Schassberger [1970] showed that a sequence of mixed Erlangs can be found which will converge weakly (that is, it converges at every point of continuity) to any arbitrary distribution function. In the sense of pointwise convergence at points of continuity, we can then say that the class of mixed generalized Erlangs (MGEs) is dense in the class of all distribution functions. The denseness of this family gives an indication of the theoretical comprehensiveness of the MGEs as a practical modeling tool.

In parallel with the above developments, computational techniques related to transform methods from complex analysis have

emerged. While the interpretation of transform methods is not always clear, the methods frequently reduce computational effort. The Laplace transform of the exponential is, of course, the reciprocal of a first-degree polynomial. Smith [1953] considered the family of distributions whose transforms are the reciprocal of an n^{th} degree polynomial (K_n). This is a natural extension of the generalized Erlang in transform space. It further admits of the interpretation as successive exponential stages in which the exponential parameters are possibly complex.

The inverse polynomial approach was extended to the rational transform (R_n) case by Cox [1955]. He showed that the formal solution methods are still valid, even though the intermediate interpretation is a bit awkward. The exponential stages may have complex valued parameters and the "probabilities" associated with the mixture interpretation may be negative. It is worth observing that the set of distribution functions with rational transforms is dense in the set of distribution functions.

While the method of rational transforms may lead to a distribution function which is as close as desired to an arbitrary distribution function, the degree of the denominator polynomial may be quite high. From a computational viewpoint, determining all the roots to a reasonable degree of accuracy may not be possible. Hence, alternative methods with reduced computational complexity are needed.

Neuts [e.g., 1981] examined the phase-type (PH) class of distribution functions, defined as the times until absorption in a finite-dimensional Markov Chain. While this family of distributions is not as comprehensive as the R_n family, it is still dense in the space of all distribution functions. It also has the computational advantage that matrix-vector procedures in the real domain are used in lieu of transform methods.

The generalized hyperexponential (GH) family of probability distribution functions is examined in detail in this paper. The GH family has the same computational advantage over transform methods that the PH family has, namely, avoidance of complex arithmetic.

We will see later that it also has some other advantages over PH representations. The GH family of probability distribution functions is partially motivated as an extension of the notion of a mixture of exponentials, where we, as Cox did earlier, permit the intermediate mixing "probabilities" to have negative values.

Botta and Harris [1986] demonstrated the denseness of the GH family in the class of all distribution functions. The present paper addresses several additional features of the GH family which recommend it as a modeling distribution with high computational tractability. A major feature of the GH distribution function is its uniqueness of representation, which is discussed in detail in the latter part of Section 2. Section 2 also provides a complete picture of the position of the GH family relative to the other families mentioned. Six closure properties of the family which relate to the mathematical operations frequently performed in applied probability are discussed in Section 3. A final section illustrates some characteristics of the GH family which facilitate numerical inversion for the determination of its quantiles. In addition, the related problem of random variable creation is addressed, using a special acceptance-rejection algorithm.

2. Relations Among Classes of Distribution Functions

Families of probability distribution functions that find wide use as approximations to more general CDFs are defined and related to one another in this section. The more obvious relations are mentioned with the definitions, while others are presented in following sections. Several of the definitions below are stated in terms of the one-sided Laplace-Stieltjes transform of a CDF, F. This transform, F*, is defined as

$$F^*(s) = \int_0^\infty e^{-st}\, dF(t),$$

which is equivalent to the ordinary one-sided Laplace transform of a PDF, $f(t) = dF/dt$, whenever F(t) is absolutely continuous.

2.1 Definitions

K_n Class

Smith [1953] defined the class K_n to be those distribution functions whose Laplace transforms are the reciprocals of polynomials of the n^{th} degree. Not all reciprocal polynomials are transforms of CDFs. For example, the real part of each polynomial root must be negative, and while the roots may be complex, they must occur in conjugate pairs since the corresponding CDF is real. There are also additional constraints that are not so obvious. Lukacs and Szasz [1951] have shown that one of the roots with greatest real part must be real. Therefore, the simplest member of K_n having complex roots is of the form

$$F^*(s) = \frac{a(a^2 + b^2)}{(s + a)[(s + a)^2 + b^2]},$$

corresponding to the PDF

$$f(t) = ab^{-2}(a^2 + b^2)e^{-at}(1 - \cos bt) \qquad (a > 0). \qquad (2.1.1)$$

The exponential distribution belongs to K_n. Since the Laplace transform of the distribution of a sum of independent random

variables is the product of the Laplace transforms of their individual distributions, the generalized Erlang CDFs corresponding to a sum of independent, exponentially distributed random variables with distinct parameters are also in K_n. These generalized Erlangs, denoted GE, have transforms of the form

$$\prod_{i=1}^{n} \frac{\lambda_i}{s + \lambda_i} \qquad (\lambda_i > 0)$$

where $\lambda_i/(s + \lambda_i)$ is the transform of an exponential CDF having mean $1/\lambda_i$. If all the random variables are identically distributed, the resulting distribution is the (simple) Erlang of degree n, $E_n(\lambda)$, and its Laplace transform is just $\lambda^n/(s + \lambda)^n$. Therefore, we see that $E_n(\lambda) \in K_n$ and

$$GE \subset K_n \quad . \tag{2.1.2}$$

R_n Class

While K_n contains GE, it does not contain mixtures of GE CDFs, i.e., distributions of the form $\Sigma a_i F_i$ with $a_i > 0$, $\Sigma a_i = 1$ and $F_i \in$ GE. Suppose, for example, each F_i is exponential. By the linearity of the Laplace transform, the transform of $\Sigma a_i F_i$ is

$$\sum_{i=1}^{n} a_i \frac{\lambda_i}{s + \lambda_i} \quad .$$

When combined into a single fraction, a quotient of two polynomials results, the degree of the denominator being n and the degree of the numerator n - 1. This motivates the definition of R_n as the class of distributions whose transforms are rational. The index n is the degree of the denominator polynomial. Hence, the class of mixed generalized Erlang distributions, denoted by MGE, is contained in R_n. Cox [1955] points out that both the convolution and the mixture of any pair of distributions in R_n yields another distribution with rational Laplace transform. Furthermore, all distributions in R_n are continuous except for possible atoms at the origin and the corresponding density function is positive everywhere in $(0,\infty)$ except at isolated points. Finally, one sees that

$$K_n \subset R_n \quad . \tag{2.1.3}$$

PH Class

Neuts [1975, 1981] has popularized a class of distribution functions known as "phase" type, or PH, distributions. A CDF is of phase type if it can be interpreted as the time until absorption in a finite-state continuous-time Markov chain. That is, F is phase type if it can be written as

$$F(t) = 1 - \underline{\alpha} \cdot e^{Qt} \cdot \underline{e} \tag{2.1.4}$$

where Q is the generator matrix and has the form

$$Q = \begin{bmatrix} -q_{11} & q_{12} \cdots & q_{1n} \\ q_{21} & -q_{22} \cdots & q_{2n} \\ \vdots & \vdots & \vdots \\ q_{n1} & q_{n2} \cdots & -q_{nn} \end{bmatrix} \quad \begin{array}{l} (q_{ii} > 0;\ q_{ij} \geq 0,\ i \neq j; \\ -q_{ii} + \sum\limits_{\substack{j=1 \\ j \neq i}} q_{ij} \leq 0,\ i = 1,2 \cdots, n). \end{array}$$

This generator matrix corresponds to an (n + 1)-state Markov chain with absorbing state (n + 1). The vector $\underline{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ is the vector of initial state probabilities at t = 0, and the vector $\underline{e}$ is an n-dimensional column vector of all ones. The entries, q_{ij}, in the generator matrix represent the instantaneous rate of the transition from state i to state j. Each component of $e^{Qt} \cdot \underline{e}$ corresponds to a phase-type distribution that results from starting in a particular state. Therefore, (2.1.4) can be interpreted as a mixture of phase-type distributions, that is,
$F(t) = \sum \alpha_i [1 - (e^{Qt} \cdot \underline{e})_i]$.

Two examples of distribution functions with PH representations follow.

Example 2.1.1 The GE distribution of order n with parameters $\lambda_1, \lambda_2, \cdots, \lambda_n$ has the representation $\underline{\alpha} = (1,0,0,\cdots,0)$ and

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 \cdots\cdots\cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 \cdots\cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots\cdot & -\lambda_{n-1} \quad \lambda_{n-1} \\ 0 & 0 & \cdots\cdot & 0 \quad -\lambda_n \end{bmatrix} .$$

Example 2.1.2 The mixed exponential distribution

$$F(t) = \sum_{i=1}^{n} \alpha_i(1 - e^{-\lambda_i t})$$

has the representation $\underline{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ and a diagonal Q matrix with elements $-\lambda_i$.

Notice that PH representations are not unique. That is, there may exist many different generator matrices of different orders that lead to the same CDF. (We provide an example later in Section 2.5.) The problem of finding minimal representations of PH distributions (that is, where the order of Q is as small as possible) is an open question. Neuts [1981] did show that the class of PH distributions is closed under convolution and finite mixtures but not under infinite mixtures. It follows that MGE distributions are phase type, i.e.,

$$\text{MGE} \subset \text{PH}.$$

The representation (2.1.4) of a PH distribution was obtained from the distribution functions, $\underline{v}(t)$, of the individual states of the underlying Markov chain which are the solutions of

$$\frac{d\underline{v}(t)}{dt} = \underline{v}(t) \cdot Q. \qquad (2.1.5)$$

The solution to this equation is $\underline{v}(t) = \underline{v}(0)e^{Qt} = \underline{\alpha}\, e^{Qt}$. Taking the Laplace transform of (2.1.5) yields

$$sV^*(s) - \underline{v}(0) = V^*(s) \cdot Q ,$$

so that

$$V^*(s) \cdot (sI - Q) = \underline{v}(0) = \underline{\alpha}$$

or

$$V^*(s) = \underline{\alpha}\,(sI - Q)^{-1} .$$

Thus $(sI - Q)^{-1}$ is the Laplace transform of e^{Qt}, and each term in the inverse matrix of $sI - Q$ is a rational expression.

Multiplication by $\underline{\alpha}$ yields rational expressions for each component of V*(s). Therefore, the probability distribution of each state belongs to R_n as does the distribution of the time until absorption. Therefore, we see that

$$PH \subset R_n \quad . \tag{2.1.6}$$

Phase-type distributions exist which possess Laplace transforms that are not reciprocal polynomials, so that $PH \not\subset K_n$. But it is not possible for every K_n distribution to have a PH representation. Corollary 2.2.1 in Neuts [1981] proves that any non-trivial PH distribution has a corresponding density function that is strictly positive for all $t > 0$. The PDF given in (2.1.1) has a reciprocal-polynomial Laplace transform but the density function is zero wherever $\cos bt = 1$. Therefore, the corresponding distribution function is not in PH and $K_n \not\subset PH$ which implies that $R_n \not\subset PH$ and that PH is thus a proper subset of R_n. Observe that, for an arbitrary CDF, there is no easy way to determine if it is in PH. One must search for a suitable generator matrix and set of initial conditions that will yield the desired distribution.

GH Class

The generalized hyperexponential distributions are CDFs of the form

$$1 - \sum_{i=1}^{n} a_i e^{-\lambda_i t}$$

with λ_i and a_i real, $\lambda_i > 0$ and $\sum_{i=1}^{n} a_i = 1$. Unlike the usual hyperexponential distribution, we do not require that each a_i be nonnegative. This added freedom makes the GH distributions extremely versatile. Indeed, Botta and Harris [1986] derived the critical characterization that any CDF on $[0,\infty)$ can be approximated by a member of GH as closely as desired with respect to an appropriate metric.

The Laplace transform of a GH distribution is

$$\sum_{i=1}^{n} \frac{a_i \lambda_i}{s + \lambda_i}$$

so one notes that

$$GH \subset R_n \qquad (2.1.7)$$

But not all functions of the form $1 - \sum_{i=1}^{n} a_i e^{-\lambda_i t}$ with $\lambda_i > 0$ and $\sum_{i=1}^{n} a_i = 1$ are GH distributions. We do know, for example, that the CDF's monotonicity requires that $\sum_{i=1}^{n} a_i \lambda_i > 0$. Also, assuming λ_n to be the smallest of the λ_i, the corresponding coefficient a_n must be positive to insure proper asymptotic behavior as $t \to \infty$. Bartholomew [1969] derived a number of sufficient conditions for a linear combination of exponentials to be a GH distribution, but no set of conditions that are both necessary and sufficient is known. Dehon and Latouche [1982] have recently characterized the class of GH distributions by deriving a parametric equation of the boundary of the convex region constituting GH for the case $n = 3$. The geometric representation is obtained by choosing a set of basis vectors from the class of all GH distributions composed of linear combinations of three exponentials. It does not appear that the boundary equation can be easily used to determine if a candidate exponential sum is, in fact, in GH. For sums of more than three exponential terms, the boundary equation could be determined in similar fashion but would be very involved and still not of much practical use in determining membership in GH.

We next develop some additional relations among the classes K_n, R_n, GE, MGE, PH, and GH.

2.2 GH and PH

From the preceding, we know that all PH distributions are in R_n. From the discussion leading up to (2.1.6), it is clear that if the generator matrix has distinct real eigenvalues, then the corresponding PH distribution also will be in GH. But if the denominator polynomial has repeated or complex roots, the

corresponding distribution will not belong to GH. The following example displays such a PH distribution.

Example 2.2.1 Consider the 3x3 generator matrix

$$Q = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 1 & 0 & -3 \end{bmatrix}.$$

The eigenvalues of Q, which are equal to the roots of the denominator polynomial of the Laplace transform of e^{Qt}, are

$$\lambda_1 = -.2307 \; ; \; \lambda_2, \lambda_3 = -2.8846 \pm .5897 \; i$$

where $i = \sqrt{-1}$. The resulting PH distribution corresponding to an initial state vector $\underline{\alpha} = (1,0,0)$ is

$$F(t) = 1 - 1.1729 \; e^{-.2307t} - [.1729 \cos .5897t + .3868 \sin .5897t] \; e^{-2.8846t}.$$

Because of the trigonometric terms, F(t) is clearly not in GH so

$$PH \not\subset GH \quad .$$

But not every GH distribution has a PH representation. As mentioned earlier, the density function corresponding to any PH distribution is strictly positive for all $t > 0$. The following example exhibits a GH distribution that violates this condition.

Example 2.2.2 Consider the GH distribution defined by

$$F(t) = 1 - (4e^{-t} - 6e^{-2t} + 3e^{-3t})$$

with corresponding density

$$f(t) = F'(t) = 4e^{-t} - 12e^{-2t} + 9e^{-3t} \; .$$

It can easily be shown that $f(t) = 0$ for $t = \ln(3/2)$ and that $f(t) > 0$ for all other values of t. Therefore, $F(t)$ is not PH and

$$GH \not\subset PH \quad .$$

2.3 MGE and GH

Recall that the generalized Erlang (GE) distributions have Laplace transforms

$$\prod_{i=1}^{n} \frac{\lambda_i}{s + \lambda_i}$$

where the λ_i are distinct. Using a partial fraction expansion, this transform can be written as

$$\sum_{i=1}^{n} \frac{A_i}{s + \lambda_i}$$

where the A_i are real. Any mixture of such distributions has a transform of the same form. So, any mixed generalized Erlang distribution is in GH and

$$MGE \subset GH \quad . \tag{2.3.1}$$

Based upon results in Dehon and Latouche [1982], we next demonstrate the existence of GH distributions that cannot be represented as MGEs of the same order. They show that any GE distribution constructed from a subset of exponential distributions, $\{F_j\}$, can be expressed as a mixture of the GE distributions $F_1, F_{12}, \cdots, F_{12\cdots n}$ where $F_{12\cdots i}$ is the convolution of the first i exponential distributions from $\{F_j\}$. Each such distribution function can be written as

$$F_{12\cdots i}(t) = \sum_{j=1}^{i} \left[\prod_{\substack{k=1 \\ k\neq j}}^{i} \left(\frac{\lambda_k}{\lambda_k - \lambda_j} \right) \right] F_j(t) \qquad (t > 0), \tag{2.3.2}$$

where $F_j(t) = 1 - e^{-\lambda_j t}$ and it has been assumed without loss of generality that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. Since the $\{\lambda_j\}$ are constants, (2.3.2) is in the form of a GH distribution whose coefficients are determined by the $\{\lambda_j\}$, which agrees with (2.3.1). In order for a

GH distribution, $F(t) = 1 - \sum a_i e^{-\lambda_i t}$, to have a MGE representation, there must exist a set of nonnegative numbers $\{b_i, i = 1,2,\cdots,n\}$ which sum to one and satisfy the equation

$$1 - \sum_{i=1}^{n} a_i e^{-\lambda_i t} = \sum_{i=1}^{n} b_i F_{12\cdots i}(t) . \qquad (2.3.3)$$

By substituting (2.3.2) into (2.3.3), collecting like terms and then equating coefficients of each exponential term on the left and right sides of (2.3.3), the following triangular system of linear equations relating the $\{a_i\}$ and $\{b_i\}$ results:

$$a_1 = b_1 + \sum_{k=2}^{n} b_k \prod_{j=2}^{k} \frac{\lambda_j}{\lambda_j - \lambda_1} ,$$

$$a_i = \sum_{k=i}^{n} b_k \prod_{\substack{j=1 \\ j\neq i}}^{k} \frac{\lambda_j}{\lambda_j - \lambda_i} \quad (i=2,3,\cdots,n). \qquad (2.3.4)$$

This system of equations is readily inverted to yield the $\{b_i\}$ in terms of the $\{a_i\}$ as follows:

$$b_1 = \sum_{k=1}^{n} \frac{\lambda_k}{\lambda_1} a_k = (1/\lambda_1) \sum_{k=1}^{n} \lambda_k a_k ,$$

$$b_i = \sum_{k=i}^{n} \lambda_k a_k \frac{\prod_{j=1}^{i-1} (\lambda_j - \lambda_k)}{\prod_{j=1}^{i} \lambda_j} \quad (i=2,3,\cdots,n). \qquad (2.3.5)$$

For the case $n = 3$, the above system of equations becomes

$$a_1 = b_1 + \frac{\lambda_2}{\lambda_2 - \lambda_1} b_2 + \frac{\lambda_2\lambda_3}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)} b_3 ,$$

$$a_2 = \frac{\lambda_1}{\lambda_1 - \lambda_2} b_2 + \frac{\lambda_1\lambda_3}{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2)} b_3 , \qquad (2.3.6)$$

$$a_3 = \frac{\lambda_1\lambda_2}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)} b_3 ,$$

and thus

$$b_1 = a_1 + \frac{\lambda_2}{\lambda_1} a_2 + \frac{\lambda_3}{\lambda_1} a_3$$

$$b_2 = \frac{\lambda_1 - \lambda_2}{\lambda_1} a_2 + \frac{\lambda_3(\lambda_1 - \lambda_3)}{\lambda_1\lambda_2} a_3 \quad . \qquad (2.3.7)$$

$$b_3 = \frac{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)}{\lambda_1\lambda_2} a_3$$

From (2.3.7) b_3 is guaranteed to be nonnegative since $a_3 > 0$ for $F(t)$ to be a distribution function and the λ_i are positive with $\lambda_1 > \lambda_2 \cdots > \lambda_n$. The nonnegativity of b_1 follows from noting that b_1 can be written as $(1/\lambda_1)\, F'(0)$ where $F'(t) = \frac{d}{dt} F(t)$. Since $F'(t)$ is the PDF corresponding to $F(t)$, it must be nonnegative for all t. Requiring that $b_2 > 0$ leads to the condition

$$a_2 > - \frac{\lambda_3(\lambda_1 - \lambda_3)}{\lambda_2(\lambda_1 - \lambda_2)} a_3 . \qquad (2.3.8)$$

The next example demonstrates that there exist GH distributions for which condition (2.3.8) is violated.

Example 2.3.1 Consider the GH CDF

$$F(t) = 1 - (6e^{-4t} - 13e^{-3t} + 8e^{-2t}).$$

Here

$$a_1 = 6, \; a_2 = -13, \; a_3 = 8$$

and

$$\lambda_1 = 4, \; \lambda_2 = 3, \; \lambda_3 = 2.$$

Therefore

$$-\frac{\lambda_3(\lambda_1 - \lambda_3)}{\lambda_2(\lambda_1 - \lambda_2)}\, a_3 = -\frac{32}{3}\,.$$

Since $a_2 < -32/3$, we see that (2.3.8) is violated and thus that no MGE representation exists for F(t). This example establishes that

$$GH \not\subset MGE\,,$$

and that the class of MGE distributions is thus a proper subset of the class of GH distributions. However, it is sometimes possible to obtain a MGE representation by embedding the problem in a higher-order space even when there is no valid MGE representation in the original space. An illustration is provided later in Section 2.5 as Example 2.5.1.

2.4 MGE and PH

We established in Section 2.1 that all MGE distributions are phase type. Since PH distributions may include trigonometric terms, it is clear that the MGE distributions are a proper subset of PH. But what if the PH generator matrix is allowed to have only real eigenvalues? Is the resulting subclass of PH distributions contained in MGE? The answer is no. We obtain this result by way of a counterexample.

Example 2.4.1 The PH distribution given by

$$F(t) = 1 - (1.293\, e^{-4.846t} - .343\, e^{-4.195t} + .050\, e^{-.959t}), \tag{2.4.1}$$

was obtained from the generator matrix

$$Q = \begin{bmatrix} -5 & 0 & 1/8 \\ 4 & -4 & 0 \\ 0 & 1 & -1 \end{bmatrix} \tag{2.4.2}$$

with $\underline{\alpha} = (1,0,0)$. As before, equating F(t) to $b_1F_1(t) + b_2F_{12}(t) + b_3F_{123}(t)$ and solving for the $\{b_i\}$ yields the result that $b_2 =$

-.0369. Since each b_i must be nonnegative, we do not have a valid MGE representation. Thus, PH distributions with real roots do not necessarily belong to MGE. In other words,

$$\text{PH (real roots)} \not\subset \text{MGE}$$

and MGE is a proper subset of PH (real roots).

2.5 Uniqueness of Representation

For statistical applications, an important property of mixture-type CDFs is uniqueness of representation, or identifiability. Yakowitz and Spragins [1968] define the identifiability of finite mixtures as follows. If $\{F_i\}$ is a collection of CDFs, then the class of finite mixtures of the $\{F_i\}$ is said to be identifiable if the convex hull of $\{F_i\}$ has the property that

$$\sum_{i=1}^{N} c_i F_i = \sum_{i=1}^{M} c_i' F_i'$$

where $c_i > 0$, $\sum c_i = 1$, implies $N = M$ and that for each i $(1 \leq i \leq N)$ there is some j $(1 \leq j \leq N)$ such that $c_i = c_j'$ and $F_i = F_j'$. A necessary and sufficient condition for identifiability is that the class $\{F_i\}$ be a linearly independent set over the field of real numbers. This follows from the uniqueness of representation property of a basis in a vector space.

Since any collection of distinct exponentials is linearly independent, the class of finite mixtures of exponential CDFs is identifiable. A broader concept of identifiability for generalized mixtures also applies when the underlying family of CDFs is exponential. A generalized mixture is one where the mixing parameters sum to unity but can have any real values; the GH distributions are of this form. Again, the uniqueness of the representation of vectors with respect to a basis for the vector space implies that GH distributions have unique representations as linear combinations of exponentials.

Importantly, the other families of CDFs considered in this work do not share the uniqueness of representation property with the GH distributions. For example, consider the following two distinct phase-type representations:

$$Q = \begin{bmatrix} -3 & 1 & 1 \\ 1 & -4 & 2 \\ 1 & 0 & -6 \end{bmatrix}, \qquad \underline{\alpha} = (0,1/2,1/2)$$

and

$$Q' = \begin{bmatrix} -2 & 0 \\ 0 & -5 \end{bmatrix}, \qquad \underline{\alpha}' = (2/3,1/3) \quad .$$

Clearly the two representations are different and are not of the same order. However, each results in the same CDF, namely, $F(t) = 1 - (2/3)\, e^{-2t} - (1/3)\, e^{-5t}$. The second representation is of minimal order since the CDF is a mixture of two exponentials.

Mixed generalized Erlang distributions also permit multiple representations. From the notation of Dehon and Latouche [1982] we may represent the CDF of the sum of n independent random variables, each exponentially distributed with parameter λ_i $(i = 1,2,\cdots,n)$, by $F_{12\cdots n}$. Now consider the two CDFs defined by

$$F(t) = (1/3)\, F_1 + (2/3)\, F_{13}$$

and

$$G(t) = (1/3)\, F_1 + (4/9)\, F_{12} + (2/9)\, F_{123} \; .$$

That these two CDFs are, in fact, the same can be seen by expressing each as a linear combination of the underlying exponential distributions. The following unique representation is obtained:

$$F(t) = G(t) = (-1/3)\, F_1 + (4/3)\, F_3 \; .$$

As in the PH example, one of the MGE representations is not of minimal order.

For most applications, such as curve fitting, non-uniqueness of representation is a disadvantage. But obtaining a representation of non-minimal order sometimes may be useful. For example, suppose we have a GH distribution that does not have an MGE representation of minimal order. It may be possible to embed the distribution in a higher-order space in such a way that an MGE representation is obtained. We illustrate the procedure via an example.

Example 2.5.1 Consider the GH distribution

$$F(t) = 1 - \left(-\frac{13}{15}e^{-7t} + \frac{77}{12}e^{-4t} - \frac{35}{4}e^{-3t} + \frac{21}{5}e^{-2t}\right) .$$

Here $\lambda_1 = 7$, $\lambda_2 = 4$, $\lambda_3 = 3$, $\lambda_4 = 2$. Dehon and Latouche [1982] established that an MGE representation exists if, and only if, there exists a set of coefficients $\{b_i,\ i = 1,2,3,4\}$ such that

$$F(t) = b_1F_1 + b_2F_{12} + b_3F_{123} + b_4F_{1234}$$

with the $\{b_i\}$ nonnegative and summing to one. It can be shown that such a set of coefficients does not exist (b_3 is negative). Let us now add an additional exponential term, e^{-6t}, and write

$$F(t) = 1 - \left(-\frac{13}{15}e^{-7t} + 0\ e^{-6t} + \frac{77}{12}e^{-4t} - \frac{35}{4}e^{-3t} + \frac{21}{5}e^{-2t}\right).$$

Here, $\lambda_1' = 7$, $\lambda_2' = 6$, $\lambda_3' = 4$, $\lambda_4' = 3$, $\lambda_5' = 2$. We must now solve for the coefficients $\{b_i'\}$ from

$$F(t) = b_1'F_1' + b_2'F_{12}' + b_3'F_{123}' + b_4'F_{1234}' + b_5'F_{12345}'$$

where the primes indicate that the corresponding terms are defined with respect to the $\{\lambda_i'\}$. It turns out that there is a solution for the $\{b_i'\}$ that results in the representation

$$F(t) = \frac{1}{4}F_1' + \frac{1}{3}F_{12}' + \frac{1}{24}F_{123}' + \frac{1}{24}F_{1234}' + \frac{1}{3}F_{12345}' \quad .$$

Not only does this give us an MGE representation, it also confirms that the original F(t) is, in fact, a valid CDF since it can be expressed as a mixture of CDFs.

Since all MGEs are of phase type and there exist GHs that are not members of PH (see Section 2.2), it is not possible to obtain an MGE representation for every GH distribution. A more complete discussion of the representation of GH distributions as MGEs, including a set of necessary and sufficient conditions that does not require solving for the $\{b_i\}$ coefficients, is contained in Botta [1986]. The uniqueness property provides a strong rationale for our interest in the GH class of distributions. It is unfortunate that the PH family of distributions does not have a corresponding uniqueness property.

2.6 Summary of Set Inclusion Relations

The results of the foregoing sections yield the following set of relations among the classes of distribution functions:

(1) $GE \subset K_n \subset R_n$

(2) $GE \subset MGE \subset GH \subset R_n$

(3) $GE \subset MGE \subset PH \subset R_n$

(4) $PH \not\subset K_n$; $K_n \not\subset PH \Longrightarrow R_n \not\subset PH$

(5) $PH \not\subset GH$; $GH \not\subset PH$

(6) $GH \not\subset MGE$ (of same order)

These relations are depicted in the Venn diagram below.

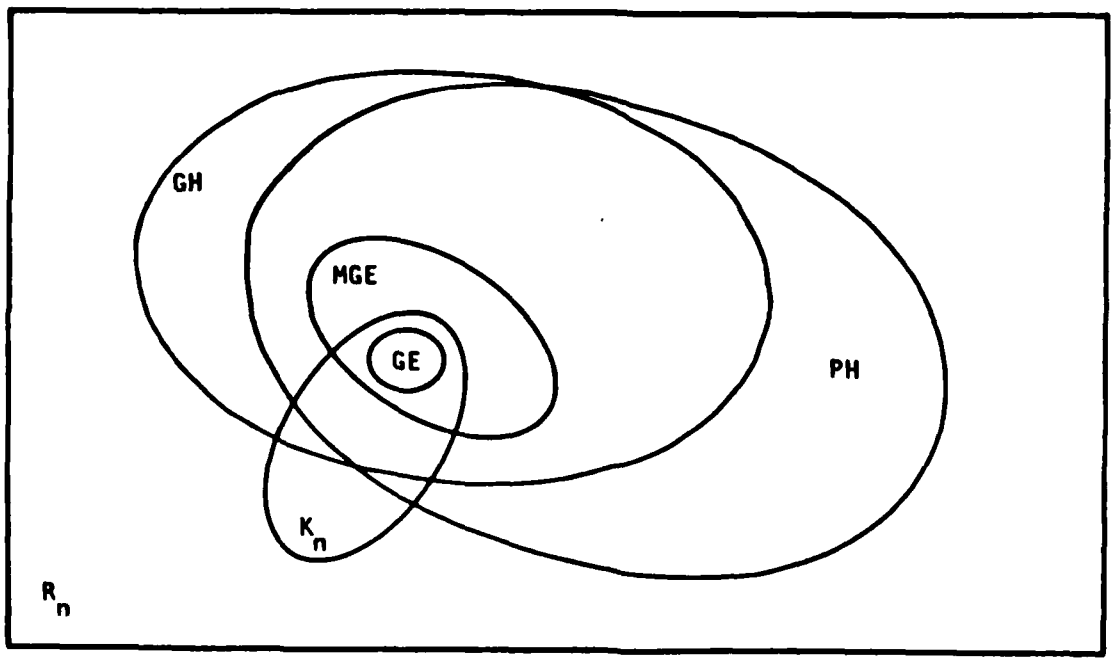


Figure 1: Set Inclusion Relations Between Classes of CDFs.

3. Closure Properties

Closure properties guarantee that the result of performing certain mathematical operations on members of a class will be another member of the same class. For example, the set of positive integers is closed under addition and multiplication. Closure is useful in applications because the outputs of certain processes will share the properties of the inputs. In this section we present some closure properties of the GH class of distributions arising in probability modeling, order statistics, reliability theory, and queueing theory. These properties provide additional justification for using GH distributions as approximations to arbitrary CDFs.

3.1 Probability Modeling

In probability modeling, mixtures of distributions and convolutions of distributions frequently arise. Neuts [1975, 1981] has shown that the PH class is closed under convolutions as well as under finite mixtures. The PH class is also closed under infinite mixtures in the sense that an infinite mixture of successive convolutions of a PH distribution, where the mixing parameters constitute a discrete PH distribution, is itself PH. Since the GH class consists of linear combinations of a finite number of exponential terms, it follows immediately that GH is closed under finite mixtures. An infinite mixture could involve an infinite number of exponential terms and would then not be in GH, by definition. Therefore, GH is closed under infinite mixtures only when the number of distinct exponential terms contained in the mixture is finite.

The GH class is not closed under convolutions. This can easily be demonstrated by noting that the GH density $f(t) = \lambda e^{-\lambda t}$ convolved with itself yields the Erlang-2 density

$$f(t)*f(t) = \lambda^2 te^{-\lambda t} \qquad (3.1.1)$$

which is not in GH. While undesirable, this lack of closure under convolution is not of much practical consequence, however. By virtue of the denseness results, we can closely approximate any Erlang distribution with a GH distribution.

The GH class of distributions is closed under multiplication. If T_1 and T_2 are two random variables with corresponding CDFs $F_1(t)$ and $F_2(t)$, then it is well-known that the distribution of $\max(T_1,T_2)$ is $F_1(t) \cdot F_2(t)$. When F_1 and F_2 are each GH, it follows that

$$\begin{aligned} F_1(t) \cdot F_2(t) &= (1 - \sum_i a_i e^{-\lambda_i t}) \cdot (1 - \sum_j b_j e^{-\alpha_j t}) \\ &= 1 - \sum_i a_i e^{-\lambda_i t} - \sum_j b_j e^{-\alpha_j t} + \sum_i \sum_j a_i b_j e^{-(\lambda_i + \alpha_j)t}, \end{aligned} \tag{3.1.2}$$

which clearly is also GH.

3.2 Order Statistics

If $t_1, t_2, \cdots, t_n$ represent the values obtained by taking n independent samples from the same distribution, the corresponding order statistics, denoted $t_{(1)}, t_{(2)}, \cdots, t_{(n)}$, are obtained by placing the samples in ascending numerical sequence; that is, $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(n)}$. In applications it is frequently required to calculate the distribution of the maximum or minimum of the sample. Since the probability that the maximum value does not exceed some value, say t, is equal to the probability that the value of each random variable $T_1, T_2, \cdots, T_n$ does not exceed t we have

$$F^*(t) = \Pr\{\max (T_1, T_2, \cdots, T_n) \leq t\} = \prod_{i=1}^{n} F_i(t) = [F(t)]^n \tag{3.2.1}$$

where F^* is the distribution of the maximum of the T_i and $F_i = F$ is the distribution function of the i^{th} sample. It follows by repeated application of (3.1.2) that $F^*(t)$ is also GH, so that the n^{th} order statistic corresponding to a sample of n observations drawn from a GH distribution is also GH. From (3.2.1) note that

this result holds even if the T_i are not identically distributed as long as each one has a GH distribution.

The distribution of the k^{th} order statistic, $1 \leq k \leq n$, is well-known (see, for example, David [1981] or Guttman et al. [1982]) and is given by

$$F_k(t) = \Pr\{T_{(n)} \leq t\} = \sum_{i=k}^{n} \binom{n}{i} F^i(t)(1-F(t))^{n-i}.$$

Using (3.1.2) repeatedly, it follows that the class GH is closed under the k^{th} order statistic for $1 \leq k \leq n$. This is so because the above summation always contains the term $F^n(t)$ which ensures that unity will be a term in the expanded expression for F_k.

GH is also closed under certain functions of order statistics such as the difference between pairs of them. We will illustrate for the special case, $X_{(n)} - X_{(1)}$, called the range of the sample. The well-known expression for this CDF is

$$H(r) = \Pr\{T_{(n)} - T_{(1)} \leq r\} = n \int_0^{\infty} f(t)[F(t+r) - F(t)]^{n-1} dt .$$

For F in GH, the integrand consists of a linear combination of exponentials so that H will also be in GH.

3.3 Reliability Theory

In reliability theory, one wishes to compute the overall reliability of a system in terms of the reliabilities of its component parts. Of course, the system reliability also depends upon the way in which the components are connected. For example, a series connection of components will fail when the first component failure occurs while a parallel connection will function as long as there is at least one good component. In general, systems consist of complex structures having both series and parallel arrangements of components so that finding the system reliability is difficult.

For each component in a system we define a binary function

$$x_i = \begin{cases} 1, & \text{if component } i \text{ is functioning} \\ 0, & \text{otherwise} \end{cases} .$$

For a system consisting of n independent components we then define a binary variable, ϕ, as

$$\phi(x_1, x_2, \cdots, x_n) = \begin{cases} 1, & \text{if the system is functioning} \\ 0, & \text{otherwise} \end{cases}$$

and ϕ is called the structure function of the system. A component, i, is said to be irrelevant if

$$\phi(x_1, \cdots, x_{i-1}, 1, x_{i+1}, \cdots, x_n) = \phi(x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_n)$$

for all combinations of the variables x_j, $j \neq i$. It is intuitively reasonable to expect a system to contain only relevant components and for improvements in the components to result in equal or better performance of the system. These ideas are formalized in the following definition of a coherent system (see Barlow and Proschan [1981]).

Definition: A system of components is said to be coherent if

a) its structure function is monotonic, i.e.,

$$\phi(x_1, \ldots, x_n) \geq \phi(y_1, \ldots, y_n) \text{ for } x_i \geq y_i \qquad (3.3.1)$$

and

b) it contains no irrelevant components.

Now suppose that each component has an associated lifetime distribution that gives the probability of the component functioning at time t. That is, let $F_i(t) = \Pr\{\text{component } i \text{ fails prior to } t\}$ so that $1 - F_i(t) = \Pr\{\text{component } i \text{ is functioning at time } t\}$. We will now show that the GH class of CDFs is closed under the formation of coherent structures. That is, if each component has a GH distribution of lifetime then any coherent structure formed from those components will also be characterized by a GH distribution of lifetime. Assaf and Levikson [1982] have shown a similar result for phase-type distributions. Because of their simple structure, it is easier to demonstrate this result for GH distributions.

Let $S = \{(x_1,\cdots,x_n) \mid x_i = 0 \text{ or } x_i = 1,\ i = 1,2,\cdots,n\}$ be the state space of the system. We denote the 2^n states in S by s_1, $s_2,\cdots, S_2n$. By the law of total probability

$$\Pr\{\text{system is functioning}\} = \Pr\{\phi(x_1,\cdots,x_n) = 1\}$$
$$= \sum_{s_i \varepsilon\ S} \Pr\{\phi = 1 | s_i\} \cdot \Pr\{s_i\}\ . \qquad (3.3.2)$$

For each state, the value of $\phi(s_i)$ is known since the state determines whether the system is functioning or not. The corresponding probabilities are therefore either zero or one. That is, $\Pr\{\phi = 1|s_i\}$ is zero or one. Since the components are assumed to be independent, $\Pr\{s_i\}$ is just the product of the probabilities that each component is functioning or not, depending upon whether the corresponding value of x_j is one or zero for the state in question. For example, in a three component system with $s_2 = (1,0,1)$, we have $\Pr\{s_2\} = (1 - F_1(t))(F_2(t))(1 - F_3(t))$. For a coherent system, $\phi(0,\cdots,0) = 0$, for otherwise $\phi \equiv 1$ and all components would be irrelevant. The probability of the corresponding state, $P(0,0,\cdots,0) = \prod_{i=1}^{n} F_i(t)$ will therefore not appear in expression (3.3.2) which will consist of a sum of products of the form $\prod_i [1 - F_i(t)] \cdot \prod_j F_j(t)$ where i ranges over all values for which $x_i = 1$ and j ranges over all values for which $x_j = 0$. Since the component distributions are GH of the form $F_i(t) = 1 - \sum_j a_{ij} e^{-\lambda_{ij} t}$, it follows that (3.3.2) will be a linear combination of exponentials, so that the lifetime CDF of the system will also be GH. We illustrate with a two-component system.

Example 3.3

Let $F_1(t) = 1 - (1/3)e^{-t} - (2/3)e^{-2t}$ and $F_2(t) = 1 - 2e^{-t} + e^{-2t}$.

We therefore obtain from (3.3.2),

$$\begin{aligned}\Pr\{\text{System is functioning}\} &= \phi(1,1)(1-F_1)(1-F_2) + \phi(1,0)(1-F_1)(F_2) \\ &\quad + \phi(0,1)(F_1)(1-F_2) \\ &= \phi(1,1)[(2/3)e^{-2t} + e^{-3t} - (2/3)e^{-4t}] \quad (3.3.3) \\ &\quad + \phi(1,0)[(1/3)e^{-t} - e^{-3t} + (2/3)e^{-4t}] \\ &\quad + \phi(0,1)[2e^{-t} - (5/3)e^{-2t} - e^{-3t} + (2/3)e^{-4t}].\end{aligned}$$

For a coherent system, $\phi(1,1) = 1$. There are two possibilities: the components are connected in series so that $\phi(1,0) = \phi(0,1) = 0$ or they are connected in parallel with $\phi(1,0) = \phi(0,1) = 1$. The corresponding results from (3.3.3) are:

a) series connection:

$\Pr\{\text{system functions}\} = (2/3)e^{-2t} + e^{-3t} - (2/3)e^{-4t}$

so that the lifetime distribution of the system is

$$F_S(t) = 1 - (2/3)e^{-2t} - e^{-3t} + (2/3)e^{-4t}.$$

b) parallel connection:

$$F_S(t) = 1 - (7/3)e^{-t} + e^{-2t} + e^{-3t} - (2/3)e^{-4t}.$$

The closure result derived above applies to any coherent structure however complex and includes arbitrary series-parallel arrangements such as k out of n systems as well as bridge structures.

3.4 Queueing Theory

The M/G/1 queue is characterized by exponential interarrival and general service-time distributions. Here we will examine the nature of the steady-state residual service times and queueing times when the service time distribution is GH.

The residual service time is the remaining service time of the customer in service at the instant a new customer arrives and the waiting time (i.e., queueing time) is the time an arriving customer must wait before receiving service. Denoting the GH service time

distribution by G(t), it follows from renewal theory (see Ross [1970]) that the distribution of the residual service time is given by

$$R(t) = \mu \int_0^t [1 - G(x)]\, dx \qquad (3.4.1)$$

where $1/\mu$ is the mean of G(t). Since G is GH, this integral is of the form

$$\mu \int_0^t \left(\sum a_i e^{-\lambda_i x}\right) dx = \mu \sum a_i \int_0^t e^{-\lambda_i x}\, dx.$$

Therefore,

$$R(t) = 1 - \sum \frac{\mu a_i}{\lambda_i} e^{-\lambda_i t} = 1 - \sum b_i e^{-\lambda_i t}$$

so that R(t) is also GH.

A simple relationship exists between the Laplace transforms of R(t) and the waiting time distribution, W(t). Denoting the Laplace-Stieltjes transform of F(t) by F*(s), Gross and Harris [1985] have shown that

$$W^*(s) = \frac{1 - \rho}{1 - \rho R^*(s)} \qquad (3.4.2)$$

where $\rho = \lambda/\mu$ is the ratio of the average arrival and service rates. By expanding the right-hand side of (3.4.2) in a geometric series and taking inverse transforms it follows that

$$W(t) = (1 - \rho) \sum_0^\infty \rho^n [R^{(n)}(t)] \qquad (3.4.3)$$

where $R^{(n)}(t)$ denotes the n-fold convolution of R(t) with itself. Since we have shown that GH is not closed under convolution, it follows that (3.4.3) need not be GH. We will demonstrate this fact with an example.

Example 3.4

Let $G(t) = 1 - 3e^{-t} + 3e^{-2t} - e^{-3t}$ with mean $1/\mu = 11/6$. This has Laplace-Stieltjes transform

$$G^*(s) = \frac{6}{(s + 1)(s + 2)(s + 3)} .$$

Corresponding to this,

$$R^*(s) = \frac{\mu\ [1 - G^*(s)]}{s} = \frac{\mu(s + 6s + 11)}{(s + 1)(s + 2)(s + 3)}$$

which is seen to be GH by expanding in partial fractions. Computing $W^*(s)$ using (3.4.2) yields

$$W^*(s) = \frac{(1 - \rho)\ s(s + 1)(s + 2)(s + 3)}{(s + 1)(s + 2)(s + 3)(s - \lambda) + 6\lambda} . \qquad (3.4.4)$$

To have a steady-state queue, λ must be less than μ (6/11 in this case). Letting $\lambda = 1/2$, the denominator of (3.4.4) becomes

$$s(s^3 + 11/2\ s^2 + 8s + 1/2).$$

The cubic expression in parentheses has a real root at approximately -.0654 so this expression can be written in factored form as

$$s(s + .0654)(s^2 + 5.4346\ s + 7.6445).$$

The discriminant of the quadratic term is less than zero, so that the quadratic has complex roots. Therefore $W^*(s)$ has complex roots which implies $W(t)$ is not in GH, although from (3.4.4) it is clearly in R_3. We note from (3.4.3) and the fact that $\rho < 1$ for equilibrium that $W(t)$ consists effectively of a finite sum of exponential and Erlang terms. Since the Erlangs can be well approximated by GH distributions, from a practical viewpoint the waiting time distribution can be treated as GH even though, strictly speaking, we have seen that it is not.

By way of comparison, we note that Neuts [1975, 1981] has shown that if the service-time distribution is phase-type then both $R(t)$ and $W(t)$ are also phase-type.

4. Numerical Inversion and Random Variate Creation

In applications it is often necessary to invert a probability distribution function; that is, given a value for F(t), find the corresponding value of t. For all but the simplest CDFs an explicit inverse cannot be found. One must then resort to numerical techniques, which may be particularly complicated when a CDF type cannot be transformed back to a single standard form. A related (and occasionally identical) problem is the derivation of random variates following a specific distribution. The inversion can itself be used, but, as is typical, there is a more efficient way to generate the variates by using the distribution's properties in a more direct way.

4.1 Inversion

Many techniques exist for solving for the (unique) root of F(t)=p, including the bisection, false position, secant, and Newton-Raphson methods. These and other techniques may be found in standard works on numerical analysis such as Conte and de Boor [1980], King [1984], Ralston [1965] and Hamming [1971]. The techniques differ in their rates of convergence, complexity or computational efficiency, and the local or global nature of their convergence. In practice, tradeoffs must be made between these characteristics in order to choose the method most appropriate for the application at hand.

All of the methods listed above are general in that they can be used to find the roots of any function. We now propose a method for inverting GH CDFs based upon their underlying structure, that is guaranteed to converge, is simple to implement, and is free of the drawbacks of some of the standard techniques. Its chief disadvantage appears to be that convergence is only linear. We will describe the method first, and then briefly compare its features with those of the other techniques.

Our method belongs to the class of so-called "fixed-point" iterations of the form $x_{n+1} = g(x_n)$. We begin with the CDF

$$F(t) = 1 - \sum a_i e^{-\lambda_i t} + \sum b_j e^{-\alpha_j t}, \quad \text{where} \quad a_i, \lambda_i, b_j, \alpha_j > 0.$$

Given a particular value F(T) we wish to determine T. We first define two additional functions,

$$G(t) = F(T) + \sum a_i e^{-\lambda_i t}$$

and (4.1)

$$H(t) = 1 + \sum b_j e^{-\alpha_j t} .$$

Therefore G(t) - H(t) = F(T) - F(t). The desired value, T, is then the abscissa of the intersection of the curves G(t) and H(t). Since F(t) is monotonic, there is only one such intersection. Furthermore, G and H are each monotonically decreasing and have the following properties:

$$G(t) > H(t) \text{ for } t < T \tag{4.2}$$

and

$$G'(t) - H'(t) = -F'(t) \leq 0 \text{ for all } t. \tag{4.3}$$

The second of these relations follows from the nonnegativity of the PDF. Since G' and H' are negative everywhere, (4.3) implies that

$$|G'(t)| > |H'(t)| \text{ for all } t. \tag{4.4}$$

The proposed technique for finding T is most easily explained by referring to the graphs of G and H. These are shown in Figure 2.

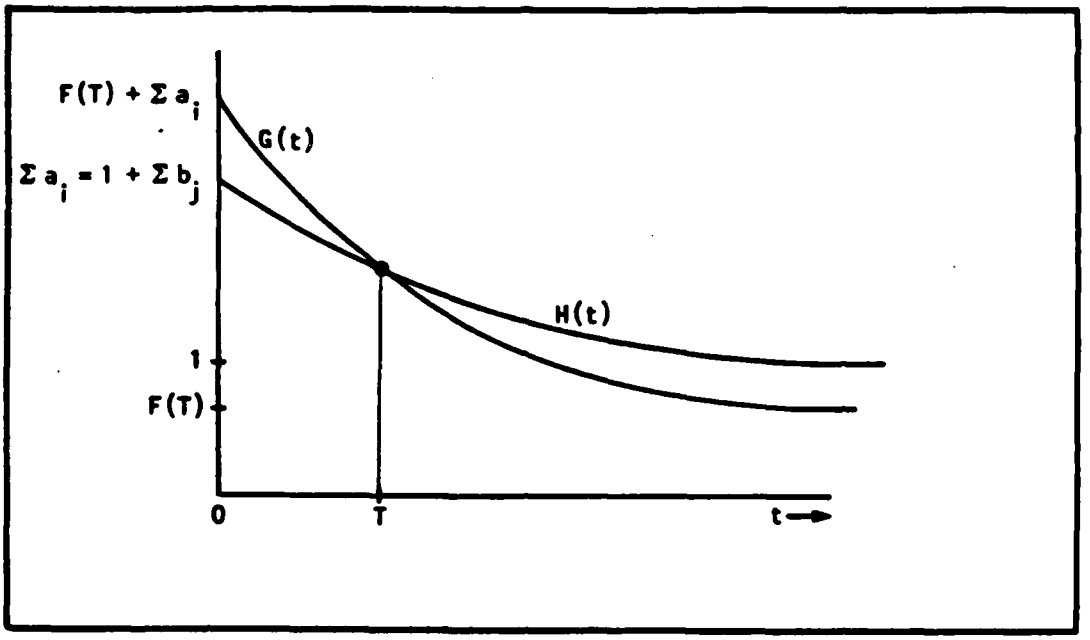


Figure 2: Geometry for Finding the Inverse of F.

Starting from any point $t_n < T$, if t is incremented in such a way that G remains greater than H, we will still be to the left of T. Repeating this process, we will converge to T from below. We next explicitly describe the procedure for incrementing t and prove convergence.

For any $t_n < T$, construct the tangent to G through the point $(t_n, G(t_n))$. By the convexity of G ($G'' > 0$ everywhere), this tangent line is always below the graph of G. Construct the horizontal line through the point $(t_n, H(t_n))$. By the monotonicity of H, this line lies above H for all $t > t_n$. Since $G' < 0$ and $G(t_n) > H(t_n)$, the tangent to G and the horizontal line intersect at a point whose abscissa is greater than t_n. Let this value be t_{n+1}. Repeat the process beginning at t_{n+1}. The sequence of steps is illustrated in Figure 3.

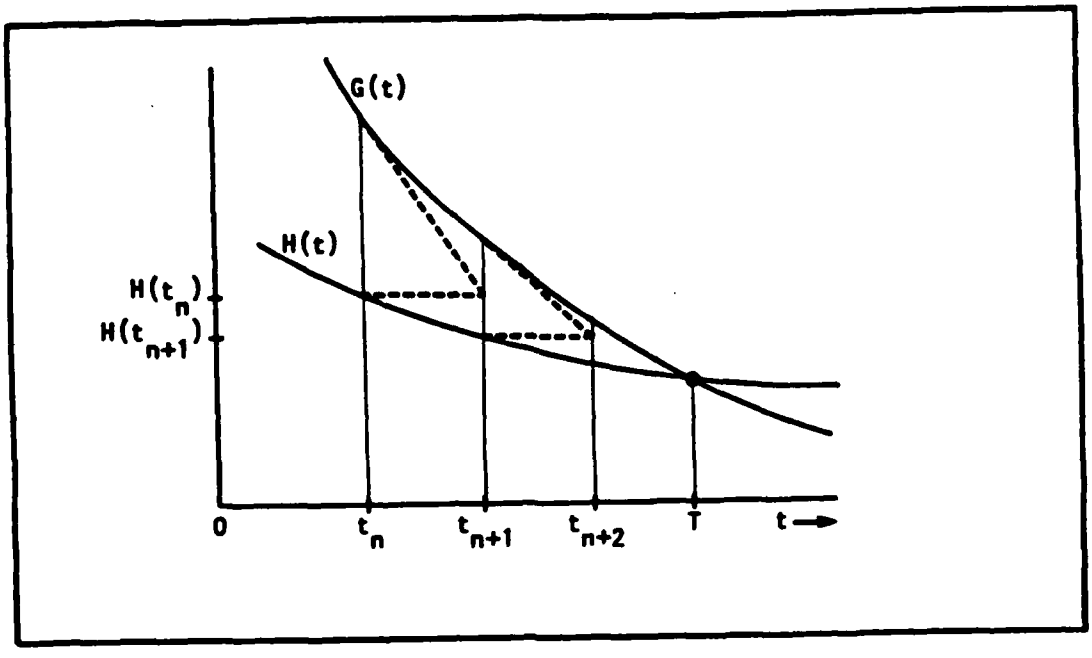


Figure 3: Iterative Scheme Illustrating Convergence.

Since the point of intersection of the tangent and horizontal lines lies on both lines we have

$$G(t_{n+1}) > H(t_n) > H(t_{n+1}). \tag{4.5}$$

From the above construction we obtain an explicit expression for t_{n+1} as follows:

$$\frac{H(t_n) - G(t_n)}{t_{n+1} - t_n} = G'(t_n)$$

or

$$t_{n+1} = t_n - \frac{G(t_n) - H(t_n)}{G'(t_n)} . \tag{4.6}$$

We next prove that the sequence $\{t_n\}$ converges to T. Since $t_n < T$, $G(t_n) - H(t_n) > 0$. From the definition of G in (4.1), $G'(t) = -\sum a_i \lambda_i e^{-\lambda_i t}$ is seen to be negative and finite for all

values of t. Therefore, from (4.6) we have that $t_{n+1} > t_n$ so that the sequence $\{t_n\}$ is monotonically increasing. From (4.5) and (4.2), $\{t_n\}$ is bounded above by T. Therefore, $\{t_n\}$ has a limit, t_o. Using this fact together with the continuity of G, H, and G', and taking the limit of (4.6) yields

$$t_o = \lim t_{n+1} = \lim t_n - \frac{\lim G(t_n) - \lim H(t_n)}{\lim G'(t_n)}$$

$$= t_o - \frac{G(\lim t_n) - H(\lim t_n)}{G'(\lim t_n)}$$

$$= t_o - \frac{G(t_o) - H(t_o)}{G'(t_o)} \quad . \qquad (4.7)$$

Therefore,

$$\frac{G(t_o) - H(t_o)}{G'(t_o)} = 0 \ ;$$

since $G'(t_o) \neq 0$ for t_o finite, this implies that $G(t_o) = H(t_o)$. Since G and H have a unique point of intersection, $t_o = T$, and the sequence $\{t_n\}$ converges to T, as desired.

Fixed-point algorithms, in general, have only linear convergence and we now demonstrate this fact for our algorithm. Denoting the error in the n^{th} iterate by $\varepsilon_n = t_n - T$, we write (4.6) as

$$\varepsilon_{n+1} = t_{n+1} - T = t_n - T - \frac{G(t_n) - H(t_n)}{G'(t_n)}$$

$$= \varepsilon_n - \frac{G(t_n) - H(t_n)}{G'(t_n)} \quad . \qquad (4.8)$$

As t_n nears T, $G'(t_n)$ is approximately equal to $G'(T)$. Using the first two terms of the Taylor series expansions for G and H about T, (4.8) becomes

$$\varepsilon_{n+1} \approx \varepsilon_n - \frac{G'(T) - H'(T)}{G'(T)} \varepsilon_n = \frac{H'(T)}{G'(T)} \varepsilon_n = K\varepsilon_n. \qquad (4.9)$$

This illustrates the linear nature of the convergence of $\{t_n\}$. From (4.4) we see that $\varepsilon_{n+1} < \varepsilon_n$ except possibly when $G'(T) = H'(T)$. In that case $F'(T) = 0$ so that the solution corresponds to a multiple root. To obtain an expression for ε_{n+1} under these conditions, more terms must be used in the Taylor series expansion in (4.8). With a triple root, we obtain

$$\varepsilon_{n+1} \approx \varepsilon_n - \frac{G'''(T) - H'''(T)}{G'(T)} \frac{\varepsilon_n^3}{3!}. \qquad (4.10)$$

Since $G(t) - H(t) = F(T) - F(t)$, $G'''(T) - H'''(T) = -F'''(T)$. At a triple root $F''(T) = F'(T) = 0$ so that $F'''(T)$ must be positive to prevent the PDF $F'(t)$ from becoming negative. As a result

$$\frac{G'''(T) - H'''(T)}{G'(T)} = C > 0$$

since both the numerator and denominator are negative. Therefore, from (4.10),

$$\varepsilon_{n+1} \approx \varepsilon_n - \frac{C}{3!} \varepsilon_n^3$$

so that the error sequence is still decreasing but at an ever diminishing rate. In a situation such as this with a multiple root, even the Newton-Raphson algorithm slows from a quadratic to a linear rate of convergence.

Our algorithm will converge to T when started at any value of $t < T$. In particular, we can take $t = 0$ initially. Like other globally convergent algorithms such as bisection or false position, convergence is linear. With the false position algorithm, however, as the solution is neared, a problem with step size may

occur since the denominator of the step expression consists of the difference of nearly equal numbers. Our algorithm does not have that problem since the denominator of the step term, G', is non-zero and only approaches zero when F(T) is very close to unity. Near the solution point, the bisection and false position algorithms can suffer from underflow since the product of two small numbers is used to determine the truncated interval containing the solution. Also, each of these algorithms requires that a pair of values bracketing the root be supplied to start the search procedure. A separate computation will usually be needed to produce such a pair of values. A slight drawback of our algorithm relative to the bisection algorithm is that the solution is approached from below so that we do not get an estimate of the error at each stage of the iteration.

Locally convergent methods, such as Newton-Raphson and secant, converge faster than our algorithm. However, convergence is not guaranteed unless the starting point is reasonably close to the solution. For example, the Newton-Raphson method may oscillate between two values and if an iteration yields a negative value of t or a zero-valued derivative, it is not clear how to proceed, short of picking a new initial point and beginning again. The secant algorithm can also produce negative values of t when both points of the current iteration are to the right of the solution and the CDF has a small slope in their vicinity. In addition, the secant method may suffer from roundoff error in the computation of step size resulting from a division by the difference of nearly equal quantities. This situation can occur when both points lie on the same side of the solution. These drawbacks offset the faster convergence of the locally convergent methods.

In practice, one often starts with a globally convergent method and switches to a more rapidly converging local method as the solution is neared. The switch to the locally convergent algorithm can be made when the step size using the initial iteration method falls below some preset threshold value. Such a hybrid scheme could be implemented by starting with our algorithm

and then switching to Newton-Raphson. This would be particularly easy since (4.6) is almost identical to the Newton-Raphson algorithm. Indeed, by simply changing the denominator to $G'(t_n) - H'(t_n)$, we obtain the Newton-Raphson iteration for finding the root of $F(T) - F(t) = 0$, which is the desired equation. Finally, we note that schemes exist for accelerating the convergence of linear-rate algorithms. One such technique is Aitken's delta-squared process (see, for example, Ralston [1965]). However, it is generally just as effective to use a hybrid scheme as described above.

4.2 The Generation of Random Variates

The decomposition of a generalized hyperexponential distribution into "positive" and "negative" functions similar to (4.1) allows an interesting application of the acceptance-rejection method to generate random variates for simulation. Our approach is a modification of the work of Bignami and de Matteis [1971], also discussed in Everitt and Hand [1981].

Let the PDF from which samples are desired be given by

$$f(x) = \sum_{i=1}^{m} p_i f_i(x) - \sum_{j=1}^{n} q_j g_j(x) \qquad (4.1.1)$$

where f_i and g_j are PDFs, $p_i, q_j > 0$ and $\sum p_i - \sum q_j = 1$. Rewrite (4.1.1) as

$$f(x) + \sum q_j g_j(x) = \sum p_i f_i(x). \qquad (4.1.2)$$

Divide through by $1 + \sum q_k = \sum p_\ell$ to obtain

$$\frac{1}{1 + \Sigma q_k} f(x) + \sum \frac{q_j}{1 + \Sigma q_k} g_j(x) = \sum \frac{p_i}{\Sigma p_\ell} f_i(x) \ . \qquad (4.1.3)$$

Both the left and right sides of (4.1.3) now are legitimate mixtures. Suppose a value of x is selected from the mixture on the right side of (4.1.3). Since this observed value is compatible with the density shown on the left of (4.1.3), it can be viewed as

arising from either f or one of the g_j. The posterior probability that it came from f can be found from Bayes' rule to be

$$\frac{f(x)}{f(x) + \sum q_j g_j(x)} = \frac{f(x)}{\sum p_i f_i(x)} . \qquad (4.1.4)$$

This suggests that if the selected value is accepted with probability given by (4.1.4), then the probability density governing the accepted value will be f(x) as desired. To summarize the procedure:

1. Generate y from the mixture $\sum (p_i/\Sigma p_\ell) f_i(y)$.
2. Generate u from U(0,1).
3. Compute $t = f(y)/(\Sigma p_i f_i(y))$.
4. If $u \leq t$ set $x = y$; otherwise discard y, return to Step 1 and repeat the process until a value of x is selected.

The discussion above is rather heuristic. We now prove in a rigorous fashion that the accepted values come from the desired distribution. Make the following definitions:

$$\sum_{i=1}^{m} p_i = 1/k ,$$

$$\sum p_i f_i(x) = w(x), \text{ and} \qquad (4.1.5)$$

$$\sum \frac{p_i}{\Sigma p_\ell} f_i(x) = kw(x) = r(x) . \qquad (4.1.6)$$

Further denote the event that a value of Y is accepted by A. We show that the distribution of the accepted values is given by $\Pr\{X \leq x\} = F(x) = \int_{-\infty}^{x} f(z)dz$. Observe that the values of X are a subset of the values of Y and that conditioned on A, values of X and Y are the same. That is

$$\Pr\{X \leq x\} = \Pr\{Y \leq x|A\} . \qquad (4.1.7)$$

By the definition of conditional probability, the right side of (4.1.7) becomes

$$\Pr\{Y \leq x|A\} = \frac{\Pr\{Y \leq x, A\}}{\Pr\{A\}} . \qquad (4.1.8)$$

Next evaluate the expressions in the numerator and denominator. From the procedure for generating Y and (4.1.6) we have that the PDF of Y is r(y). We obtain an expression for $\Pr\{A\}$ by integrating the joint distribution of A and Y over the domain of Y, as follows:

$$\Pr\{A\} = \int_{-\infty}^{\infty} \Pr\{A,y\}\, dy = \int_{-\infty}^{\infty} \Pr\{A|Y = y\}r(y)dy . \qquad (4.1.9)$$

From the acceptance procedure and (4.1.5),

$$\Pr\{A|Y = y\} = \Pr\{U \leq f(y)/w(y)\} = f(y)/w(y). \qquad (4.1.10)$$

Substituting (4.1.10) into (4.1.9) yields

$$\Pr\{A\} = \int_{-\infty}^{\infty} [f(y)/w(y)]r(y)dy = \int_{-\infty}^{\infty} kf(y)dy = k, \qquad (4.1.11)$$

recalling that r(y) = kw(y). Next evaluate the numerator of (4.1.8). Again, integrate the joint distribution over y:

$$\begin{aligned}\Pr\{Y \leq x,A\} &= \int_{-\infty}^{\infty} \Pr\{Y \leq x,A|Y = y\}r(y)dy \\ &= \int_{-\infty}^{x} \Pr\{Y \leq x,A|Y = y\}r(y)dy , \qquad (4.1.12)\end{aligned}$$

where the second equality follows from the fact that the event $\{Y \leq x\}$ has zero probability for $Y = y > x$. By the same token, the event $\{Y \leq x\}$ is certain when conditioned on $Y = y \leq x$ so (4.1.12) becomes, with the help of (4.1.10),

$$\begin{aligned}\Pr\{Y \leq x,A\} &= \int_{-\infty}^{x} \Pr\{A|Y = y\}r(y)dy \\ &= \int_{-\infty}^{x} [f(y)/w(y)]r(y)dy \qquad (4.1.13)\end{aligned}$$

$$= k \int_{-\infty}^{x} f(y)dy \quad ,$$

with the last equality following from (4.1.6). Finally, substituting (4.1.11) and (4.1.13) into (4.1.8) yields

$$\Pr\{Y \leq x|A\} = \int_{-\infty}^{x} f(y)dy \quad ,$$

and this together with (4.1.7) establishes the desired result. This proof is based on that given on p. 273 of Law and Kelton [1982] for the general case of acceptance-rejection generation of random variates distributed according to any continuous distribution.

5. Conclusion

Among the areas of future work that present themselves, an obvious and practical question is that of deciding how many terms to include in the GH mixture. This issue is related to the denseness and identifiability notions. However, no general formal procedure is found in the extant literature. Experience cited by Harris and Sykes [1986] suggests that a relatively small number of terms is often quite adequate for fitting raw data. However, definitive guidelines for determining the precise number of terms are needed.

This study has introduced the family of generalized hyperexponential (GH) distributions and demonstrated desirable properties of this class which make them attractive as approximations to general CDFs. Among the features noted in this work are:

1. The GH class is dense in the set of all CDFs defined on $[0,\infty)$ so that a GH CDF may be found that is as close as desired (with respect to a suitable metric) to any specified CDF.
2. GH distributions have a simple mathematical structure that facilitates such operations as differentiating, integrating, and taking Laplace transforms. These operations frequently occur in application areas where GH approximations can be used.
3. GH distributions have rational Laplace transforms and are therefore Coxian.
4. A GH distribution has a unique representation as a linear combination of exponential terms. This property is useful both for identification of a CDF as a member of GH and for such statistical procedures as parameter estimation.
5. Since GH CDFs are composed of exponentials, the vast literature dealing with exponential and hyperexponential distributions is relevant and techniques for these common distributions can often be adapted for use with GH distributions.

6. A computer-based maximum likelihood estimation algorithm exists for fitting a GH distribution to sample data. This provides a practical advantage for modeling with GH distributions.
7. A simple globally convergent numerical algorithm exists for inverting a GH distribution. Being able to invert a CDF is important for statistical applications, such as calculating distribution percentiles.
8. The GH class of CDFs is closed with respect to operations occurring in a diverse collection of application areas including basic probability and statistics, reliability theory, and queueing theory.

REFERENCES

Assaf, D. and Levikson, B. (1982). Closure of Phase-Type Distributions Under Operations Arising in Reliability Theory. Annals of Probability, 10, 265-269.

Barlow, R. E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing: Probability Models, New York: Holt, Rinehart and Winston.

Bartholomew, D. J. (1969). Sufficient Conditions for a Mixture of Exponentials to be a Probability Density Function. Annals of Mathematical Statistics, 40, 2183-2188.

Bignami, A. and De Matteis, A. (1971). A Note on Sampling from Combinations of Distributions. Journal of the Institute of Mathematics and Its Applications, 8, 80-81.

Botta, R. F. (1985). Stochastic Modeling with Generalized Exponential Mixtures. Ph.D. Dissertation, Department of Systems Engineering, University of Virginia, Charlottesville, VA.

Botta, R. F. and Harris, C. M. (1986). Approximation with Generalized Hyperexponential Distributions: Weak Convergence Results, to appear in Queueing Systems: Theory and Applications.

Conte, S. D. and de Boor, C. (1980). Elementary Numerical Analysis: An Algorithmic Approach, 3rd ed. New York: McGraw-Hill.

Cox, D. R. (1955). A Use of Complex Probabilities in the Theory of Stochastic Processes. Proceedings of the Cambridge Philosophical Society, 51, 313-319.

David, H. A. (1981). Order Statistics, 2nd ed. New York: Wiley.

Dehon, M. and Latouche, G. (1982). A Geometric Interpretation of the Relations Between the Exponential and Generalized Erlang Distributions. Advances in Applied Probability, 14, 885-897.

Everitt, B. S. and Hand, D. J. (1981). Finite Mixture Distributions. New York: Chapman and Hall.

Gross, D. and Harris, C. M. (1985). Fundamentals of Queueing Theory, 2nd ed. New York: Wiley.

Guttman, I., Wilks, S. S., and Hunter, J. S. (1982). Introductory Engineering Statistics, 3rd ed. New York: Wiley.

Hamming, R. W. (1971). Introduction to Applied Numerical Analysis. New York: McGraw-Hill.

Jensen, A. (1954). A Distribution Model Applicable to Economics. Copenhagen: Munksgaard.

King, J. T. (1984). Introduction to Numerical Computation. New York: McGraw-Hill.

Law, A. M. and Kelton, W. D. (1982). Simulation Modeling and Analysis. New York: McGraw-Hill.

Lukacs, E. and Szasz, O. (1951). Certain Fourier Transforms of Distributions. Canadian Journal of Mathematics, 3, 140-144.

Neuts, M. F. (1975). Probability Distributions of Phase Type. In Liber Amicorum Professor Emeritus Dr. H. Florin. Louvain, Belgium: University of Louvain, 173-206.

Neuts, M. F. (1981). Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Baltimore: Johns Hopkins University Press.

Ralston, A. (1965). A First Course in Numerical Analysis. New York: McGraw-Hill.

Ross, S. M. (1970). Applied Probability Models with Optimization Applications. San Francisco: Holden-Day.

Schassberger, R. (1970). On the Waiting Time in the Queueing System GI/G/1. Annals of Mathematical Statistics, 41, 182-187.

Smith, W. L. (1953). On the Distribution of Queueing Times. Proceedings of the Cambridge Philosophical Society, 49, 449-461.

Yakowitz, S. J. and Spragins, J. D. (1968). On the Identifiability of Finite Mixtures. Annals of Mathematical Statistics, 39, 209-214.

DISTRIBUTION LIST

Copy No.	
1	Office of Naval Research 800 North Quincy Street Arlington, VA 22217 Attention: Scientific Officer, Statistics and Probability Mathematical Sciences Division
2	ONR Resident Representative Joseph Henry Building, Room 623 2100 Pennsylvania Avenue, N.W. Washington, DC 20037
3 - 8	Director, Naval Research Laboratory Washington, DC 20375 Attention: Code 2627
9 - 20	Defense Technical Information Center Building 5, Cameron Station Alexandria, VA 22314
21 - 29	C. M. Harris
30	GMU Office of Research

DTIC

END

4-86